6

"Open and Collaborative" Research: A New Model for Biomedicine

Arti K. Rai

The advent of open source software has prompted theoretical speculation about the applicability of open source innovation principles to biomedical research. This chapter moves beyond theoretical analysis into an empirical examination of existing projects that operate under an "open and collaborative" model. Open and collaborative projects often diverge from traditional open source innovation, particularly in terms of restrictions on participation, need for public funding, and use of "copyleft" licensing. Nonetheless, they represent a fresh approach to biomedical research in that they reject its exclusionary behavior and small-lab-based structure. Open and collaborative biomedical research is a potentially valuable experiment. It has produced software and genomic data that can freely be used by follow-on innovators. The model may also allow a more coordinated and comprehensive attack than has heretofore been possible on the sorts of problems that cause promising drug candidates, particularly for complex diseases, to fail. On the other hand, if the open, collaborative experiment is going to succeed, particularly in the wet lab arena, proponents must recognize the need to retain downstream patents and either work within or attempt to change publishing norms in biological science.

 I thank Jamie Boyle, Iain Cockburn, Wesley Cohen, Bob Hahn, and participants at the April 30, 2004, AEI-Brookings workshop on Intellectual Property Rights in Frontier Industries: Software and Biotech for very helpful comments. A longer version of this paper is available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=574863.

In the last twenty-five years, biomedical research has become increasingly proprietary[1] and secretive.[2] Given the cumulative nature of research, this trend has raised fears that future progress may be impeded by access and licensing difficulties. One important response has involved calls for improving access by requiring scientists and research institutions to put data and certain types of research tools into the public domain or, at a minimum, license them widely and nonexclusively at a reasonable fee.[3] This emphasis takes the current organizational structure of research as a given but seeks to reduce the intensity of exclusionary behavior associated with the research. A more dramatic response has begun to emerge, however. Public funding bodies, prominent scientists, and even some pharmaceutical firms have taken steps in the direction of what might be called *open and collaborative*[4] science. Open and collaborative projects both disavow exclusionary behavior and move beyond the traditional small-lab-based structure of biomedical research.

The rise of arguments for open and collaborative biomedical research has coincided with two phenomena: (1) the increased importance of computation in such research;[5] and (2) the well-documented emergence of so-called open source methods of innovation in computation-heavy areas of research and development, primarily software. In some recent cases, the modeling on open source software has been quite explicit; for example, the federally funded haplotype mapping project, which aims to create a database that catalogues human genetic variation, has adopted a licensing policy that is self-consciously modeled on the "copyleft" system of open source software licensing.[6] Under the copyleft version of open source development, users can access source code freely, but such access is conditioned on the user's making his improvements to such information available under the same conditions.

Although some commentators have theorized about the application of "open source" principles to biomedical research,[7] they have not analyzed carefully how this model is actually being used. In this chapter, I use information gathered through empirical investigation of existing projects to offer observations on how the open and collaborative model actually works. The projects encompassed by this label vary quite substantially. Some closely resemble open source software while others diverge rather significantly, particularly in four respects: restrictions on participation, degree of

centralization and standardization, reliance on public funding, and use of "copyleft" licensing.

Whether the open and collaborative model is likely to promote socially desirable biomedical innovation,[8] either as an absolute matter or relative to more traditional exclusionary models, is a difficult question to answer. Because the model is quite fresh, and the time delay before research on this model can be translated into end products is long, empirical demonstration of the model's virtues and vices, at this stage, probably is impossible.

Nonetheless, there are reasons to believe that the model is worth pursuing. Not only has it produced software and genomic data that are usable, but the resulting public domain status for this software and data can reduce transaction costs and secrecy that may impede the follow-on research that leads to end products. The model's least intuitive but most exciting application may involve "wet lab" systems biology: In this context, the model may allow a more-coordinated and -comprehensive attack on large, complex problems than traditional small-lab biology. Given that the dearth of knowledge regarding systems biology appears to be an important reason many promising drug candidates, particularly for complex diseases, currently fail in preclinical or clinical trials,[9] open and collaborative approaches may be welcome news even to those industries that favor strong intellectual property protection, such as the pharmaceutical industry.[10]

Certain applications of the open and collaborative model, however, raise concerns. One concern involves the possibility of reduced incentives for development and commercialization as research moves downstream, toward the chemical compound that will be the drug candidate. As various empirical studies have documented, patents on chemical compounds are critical for recouping the large costs associated with preclinical and clinical R & D.[11] To preserve development and commercialization opportunities, there is reason to be cautious about copyleft licensing, at least outside the context of software. The first section of this chapter gives economic and institutional background on biopharmaceutical innovation, with an eye toward highlighting transaction cost and secrecy concerns to which the open and collaborative model aims to respond. The next section discusses how large-scale collaboration has operated in the context of software and some other Web-based information projects. It also discusses preliminary results from empirical investigation of some prominent open and collaborative biomedical research

projects. Then, I use these results as well as the theoretical literature to elucidate the extent to which the open and collaborative model may produce socially desirable biomedical innovation. In this section, I also make recommendations for removing institutional obstacles in those cases where the model may be superior to alternative arrangements.

The Open and Collaborative Model in Context

Innovation in Biopharmaceuticals. For much of the twentieth century, biopharmaceutical innovation largely comprised trial and error by large, vertically integrated pharmaceutical firms. Through a combination of size and monopoly-conferring end product patents, these firms hedged the risk associated with their trial-and-error-based innovation. Biomedical research science operated in a different world, linked more closely to the realm of philosophy than the realm of commerce. This is not to say that biomedical science necessarily adhered in all respects to the norms of scientific communalism described by sociologists like Robert Merton.[12] The authors of one recent study that reanalyzes data from the 1960s argue that, even in 1966, experimental biologists were more reluctant than scientists in other fields to discuss ideas freely outside their individual lab.[13] Nonetheless, the secrecy that existed was fueled by academic competition, not commercial competition.[14]

In the middle to late 1970s, the advent of recombinant DNA and monoclonal antibody research caused the conceptual gap between research science and the therapeutic products of interest to industry to shrink. Just as university research was becoming interesting to industry, Congress passed the Bayh-Dole Act of 1980, which aimed to encourage downstream commercialization by allowing patenting and exclusive licensing of federally funded discoveries.[15] The Court of Appeals for the Federal Circuit, created in 1982, further encouraged proprietary trends in basic biomedical research by relaxing the so-called utility and nonobviousness requirements for patentability.[16] Finally, over the last ten years, with the infusion of genomic and proteomic information, biopharmaceutical innovation has become perhaps even more science intensive. All pharmaceutical firms aim to produce drugs by systematically testing their drug compound libraries on genomic and proteomic "targets."

The consequences of these changes have been dramatic. Universities can, and often do, patent upstream research.[17] So do small firms and startups. In the case of universities, licensing upstream research produces revenue. For small firms and startups, upstream patents, or exclusive licenses to upstream university patents, appear to be important for attracting venture capital and securing revenues when licensed to large pharmaceutical firms. Even for research that is not patented, upstream players may leverage their physical control over the data or tool to exact reach-through royalties. For their part, large pharmaceutical firms, once vertically integrated engines of innovation, must now negotiate a complex array of university and small firm proprietary claims on research inputs. While some of these claims may be narrow in scope,[18] other claims may be broader.[19] Significantly, with the increasing importance of computation, particularly software, in biomedical research, software is now another category of patented research tool that may add to upstream complexity.

Vertical "Dis-Integration" and Calls for Access. As noted, property rights on upstream research inputs have fostered the creation of small firms that market such inputs. To the extent that small firms may be more innovative than large firms,[20] and thus produce research inputs better and faster than large firms, this change could be positive. Additionally, to the extent that research inputs are licensed widely to interested downstream developers, the creation of a market for such inputs could conceivably increase downstream competition. On the other hand, as economist Ronald Coase would predict,[21] the move away from the vertically integrated firm has increased transaction costs substantially. Although such increases do not appear to have caused ongoing projects to stop,[22] there is some evidence that firms, including small firms that aim to move downstream, may avoid research areas where there are significant patent positions.[23] Additionally, in at least some cases, patent rights on research inputs have been licensed exclusively. Although exclusivity might be appropriate when a particular input needs further investment for development and commercialization, some of these cases have involved situations where the exclusive right has been asserted to block the marketing of end products that downstream developers were able to make without the need for exclusivity.[24]

Moreover, problems associated with even purely academic access appear to have become more prevalent.[25] With respect to data and materials to which researchers need physical access, both increased commercialization and increased scientific competition among labs have contributed to access difficulties. According to a 1997 study, 20 percent of academic respondents delayed publication for more than six months, either for reasons related to commercialization (for example, the need for secrecy before a patent application is filed or the desire for trade secrecy)[26] or because of scientific competition.[27] A 2002 study found that even access denials related to published research were prevalent; in 21 percent of cases, such denials caused the academic investigator to abandon a promising line of research.[28] Once again, both commercial considerations and academic competition were cited as reasons for access denials.[29] For materials that academic researchers can readily reproduce on their own, some of these researchers have been able to secure an informal regime of price discrimination by simply ignoring patent rights. Whether this informal price discrimination regime will survive the Federal Circuit's decision in *Madey v. Duke* to eliminate any possibility of an infringement exemption for research is unclear.

Beyond Access: Open and Collaborative Research

As the web of upstream proprietary rights and secrecy has grown, various public and private sector groups have made calls for greater access to research tools. Given the problems that may be created by new proprietary rights and are almost certainly created by ever-increasing levels of secrecy, calls for access are important. Indeed, in the biological sciences, such calls for access may even create a Mertonian sphere more robust than that which existed before 1980. One step beyond such calls, however, is the open and collaborative model. This model is Mertonian in the sense that scientists work openly, without secrecy and the usual sorts of exclusionary proprietary rights. But it goes beyond Merton in that it explicitly requires scientists to work closely with others outside their own lab or small firm. Requiring scientists to work closely with others, in larger groups than is ordinarily the case, responds to a set of problems that involve not only intellectual property rights but also science policy. Specifically, because complex diseases involve multiple interactions

between multiple genes and proteins, understanding these diseases may require the coordinated work of more than a small lab or firm.

In this section, I describe a variety of existing open and collaborative biomedical research projects. Because many proponents of this research invoke the example of open source software, I first discuss briefly the open source model.

The Open Source Model. Open source software development has its origins in the norm-based Mertonian framework for conducting scientific research. More specifically, the open source movement originated in a communal "hacker" culture that prevailed in certain academic laboratories in the 1960s and 1970s. At that time, packaged software was rare and individuals freely exchanged software and underlying source code for purposes of modification and improvement. Such exchange was facilitated with the creation of the ARPANET network, which eventually expanded into the Internet. Indeed, the transaction cost-lowering properties of the Internet probably allowed Mertonian norms to operate more effectively and in larger groups than they ordinarily operate.[30]

Open source software production differs from Mertonian norms in several respects. One obvious difference is that open source software development operates under the legal framework of a copyright license: The copyright in the source code is the foundation for the license. The open source license, as defined by the Open Source Initiative (OSI), now encompasses over thirty types of copyright licenses for source code. Although the exact terms of these licenses vary and some involve fees for use, they share the requirement that the licensee receive source code and be able to redistribute the source code.[31] As a first approximation, open source licenses can be divided into two categories: "copyleft" or GPL licenses that require licensees who make improvements to the software to make those improvements publicly available on the same open source terms that they received the software,[32] and those that disclose source code but essentially impose few if any requirements on recipients. An argument often made in favor of copyleft licenses is that, by preventing private appropriation of volunteer labor, such licenses provide an incentive for volunteers to contribute in the first instance.[33] The existence of myriad licenses notwithstanding, no case involving an open source license appears to have been litigated to judgment.[34] Rather, according to one study,

the primary vehicle for enforcement is identification and critiquing of violations on online mailing lists and bulletin boards.[35] Therefore, this difference from Mertonian norms may be more apparent than real.

A difference that is perhaps more significant lies in mechanisms for information integration. While the Mertonian model does not posit a specific mechanism for information integration, open source software production, particularly production in large-scale projects, often has a central developer or group of developers responsible for evaluating and integrating developments on an ongoing basis. New and modified code deemed to be of sufficient quality by the developer may then be added to the official version of the code.[36] To some extent, the control exercised by the developer resembles that exercised by firm management. On the other hand, entry and exit from developer status are more fluid than entry and exit from firm management. Thus, particularly for large-scale projects, open source software production could be seen as lying somewhere between Mertonian norms and the firm.[37]

Perhaps the most prominent respect in which open source software production differs from Mertonian science is that it is generally *not* funded publicly.[38] To the contrary, certain firms have been built on providing services for open source software. Moreover, according to one recent study of 287 open source projects, 38 percent of open source software developers make their contributions at work, with the knowledge of their supervisors. Presumably, the firms for which these developers work value the specific improvements that the developers make. Perhaps because open source contributors are highly varied in their background, they have a wide variety of intrinsic and extrinsic motivations for contributing: personal enjoyment, sense of community obligation, pay, solving a specific problem, honing skills, and enhancing career prospects.[39] Nonetheless, many open source developers, like the idealized Mertonian scientist, appear to be significantly motivated by the personal enjoyment derived from undertaking creative tasks. To the extent that this hedonic incentive or other nonmonetary incentives substitute for salary, the result can be software production that is significantly cheaper than commercial production.

Proponents of open source software argue that such software development works in the sense that it produces usable output at a lower cost than conventional proprietary development. Some make the more ambitious claim that this output, which large numbers of independent programmers

continually examine for defects and the possibility of adding additional features, is likely to be technically superior to closed source output. A small number of technical studies have tested the latter claim. One academic study that compared Linux, Apache, and GCC with their closed-source counterparts appears to buttress claims that open source software may be technically superior. The study determined that open source software had a higher rate of function modification (i.e., fixing of defects) and added more functions over time.[40] Similarly, Reasoning, Inc., a software inspection service, determined in a 2003 report that the Linux TCP/IP stack had fewer defects than commercially developed TCP/IP stacks. Reasoning, Inc.'s analysis found, however, that Apache had as many defects as its commercial counterpart. According to the authors of the latter study, this result may be a consequence of the Apache product's still being relatively early (as compared with Linux) in the software life cycle. Given the heterogeneity of both open and closed source software, attempts to generalize from a small number of case studies are perilous. Nonetheless, these studies show that open source software is a reasonable alternative to closed source, particularly if the end user wants a low price.[41]

In recent years, various large, Web-based collaborative projects have begun to generate outputs other than software. Several technical projects merit discussion because they use mechanisms for integration of information that are less hierarchical than those used in open source software development. One of these, the NASA Clickworkers project, relies on public volunteers to mark craters on Mars. Once contributions of these workers were aggregated by focusing on areas of consensus, their contributions averaged out as comparable to that of trained geologists. The technology magazine *Slashdot* uses a complicated peer rating mechanism to determine the extent to which a contributor's posts will be seen by other users. One recent review of these collaborative projects argues that such projects are likely to be superior to firms and markets in allocating human creativity when the cost to volunteers of contribution is low and such contributions can be readily filtered and aggregated.[42]

The rise of open source software and, more generally, information production through volunteer labor aggregated over the Internet has coincided with the ascendance of computation in biological research. This coincidence has inspired speculation about the possibility that similar approaches could apply to biomedical research. It has also inspired a fair number of open and

collaborative projects, particularly in the area of bioinformatics. The remainder of the paper describes these projects and evaluates the extent to which they are likely to (1) produce usable output, (2) alleviate transaction cost and secrecy problems without causing problems for commercialization, and (3) address scientific challenges, particularly challenges surrounding systems biology, that cause drug candidates for complex diseases to fail in preclinical and clinical trials.

Open and Collaborative Biomedical Research. In this section, I describe various efforts at open and collaborative biomedical research. Because the relevant technical, organizational, and economic considerations are distinct, I treat software, databases, and "wet lab" biology as separate categories. The following section then turns to an evaluation of the projects.

Bioinformatics software. Many bioinformatics software projects, particularly small software projects, operate under an open source model. By those who participate in such projects, the open source model is seen as a good mechanism for information dissemination, reduction of duplicative effort, and rapid development of software.[43] By the same token, devotees of open source do not necessarily believe that all bioinformatics software should be open source.[44]

An important difference between most open source software and open source bioinformatics software is that the latter is publicly funded. Moreover, because most research universities require that employee rights in software developed using university resources be assigned to the university,[45] the policy of universities toward open source software development becomes quite relevant.

Preliminary results from interviews with technology transfer offices (TTOs) at twenty universities that have large software patent, biomedical patent, or biomedical research portfolios indicate that most university TTOs have not, at least thus far, been seeking many software patents.[46] The reason is economic: Because software licensing typically yields little in the way of licensing revenues, software does not generally merit the cost of a patent filing. To the extent that universities distribute software, it is through nonexclusive copyright licensing.[47]

Even though they do not typically seek patents, many universities are only beginning to formulate policies with respect to open source software.[48] Two universities that have relatively well-developed policies, the University of Washington and Georgia State, treat software differently depending on whether it is perceived as commercially valuable.[49] For software that is not commercially valuable, the researcher's preference governs. If software is commercially valuable, both universities recommend that software and source code be licensed free of charge to noncommercial users but licensed for a fee to commercial users. Of course, this differentiation between commercial and noncommercial can be maintained only through limits on redistribution of source code. Such limits are in tension with open source principles that counsel against such limits.

A few universities report "bright-line" policies regarding open source software that appear more encouraging to open source. For example, both MIT and Stanford allow different types of open source software licensing if the researcher wants to use that approach.[50] MIT also manages open source licenses for researchers. Similarly, the University of Texas defers in significant part to the licensing preferences of the researcher and also manages the researcher's licenses.[51]

Genomics database projects. The first, and probably still most important, open and collaborative genomic database project was the publicly funded project to sequence the human genome. Unlike traditional human genetics, which revolved around individual laboratories that tended to be highly competitive—and hence uneven in their willingness to share information, particularly prepublication—the Human Genome Project (HGP) was, from the outset, a collaborative endeavor. The intensity of the collaboration increased in 1998, after the project was faced with a challenge from Craig Venter, the leader of a private effort to sequence the genome. To meet this challenge, the public project streamlined the number of participants and further integrated its operations. The major sequencing centers (the so-called G-5) were required to report their progress in weekly conference calls with the funding entities, principally the National Human Genome Research Institute (NHGRI).[52]

True to the nature of the open and collaborative model, the producers of the human genome sequence did not simply put the raw data into public

domain. Rather, as the data were produced, an open source software program, known as the distributed annotation system (DAS), was set up to facilitate collaborative improvement and annotation of the genome. DAS has also been applied to other genomes, including mouse, C. elegans, fruit fly, and rice. Under the DAS system, any interested party can set up an annotation server. DAS enables end users of the information—in other words, researchers—to choose the annotations they want to view by typing in the URLs of the appropriate servers. Annotation quality is judged via mechanisms somewhat similar to those employed in the NASA Clickworkers project. Specifically, according to Lincoln Stein, one of the designers of the DAS, it was "designed to facilitate comparisons of annotations among several groups. The idea is that an annotation that is similar among multiple groups will be more reliable than an annotation that is noted by one group."[53] The quality of the annotation is also judged by looking at published papers that describe the annotation technique.

The data dissemination and improvement policy of the HGP and large-scale genome mapping projects more generally was developed by scientists and National Institutes of Health (NIH) administrators and essentially imposed on the administrators of the participating universities. Although universities played no role in formulating the policy, they appear to have acquiesced in the rejection of proprietary rights.[54] Therefore, the NIH did not need to invoke the cumbersome legal procedure set up by Bayh-Dole to restrain university patenting.

Within the HGP, there was some discussion about using some type of "copyleft" license on the data produced by the project.[55] The view among these participants was that such a license would prevent private entities, particularly Craig Venter, from gaining an advantage over the public project by making proprietary any improvements Celera made to the public data. Although the HGP leaders rejected a copyleft approach,[56] NHGRI, together with other funding organizations, quite explicitly adopted a copyleft-style policy in setting up the International Haplotype Mapping Project (HapMap). This project aims to catalog haplotypes (patterns of genetic variation) and link such patterns to disease phenotypes. To identify a particular haplotype, researchers must first identify the individual genotypic variations that make up the haplotype. The HapMap project releases individual genotype data as soon as it is identified. Before haplotype information has been assembled, it may be possible for those who access the data to take these data, combine

them with their own genotype data, and generate enough information to file patent applications on haplotypes of interest. To address this possibility, the project set up a "click-wrap" license that requires those who access the HapMap database to agree that they will not file product patent applications in cases where they have relied in part on HapMap data.[57] Although this license does not (and cannot) rely on an assertion of copyright in the underlying data, it does represent an enforceable contract.[58]

Systems biology and "wet-lab" projects. Outside the context of digital information (that is, for projects that require significant wet-lab biology[59]) the open and collaborative model does not appear to have been used widely. However, it may be making some inroads in the context of some recent "systems biology" projects funded by the NIH. In the last five years, the National Institute of General Medical Science (NIGMS) has funded five large grants intended to "make resources available for independently funded scientists to form research teams to solve a complex biological problem that is of central importance to biomedical science . . . and that would be beyond the means of any one research group." These grants depart from the traditional biological grant model, which focuses on individual laboratories.

The Alliance for Cell Signaling (AFCS) was the first of these large grants to be funded. It was inspired by the HGP,[60] and it clearly invokes significant elements of an open and collaborative approach. The alliance is led by Nobel laureate Alfred Gilman of the University of Texas, Southwestern Medical School. Gilman won the Nobel Prize for his work on the role of G proteins in cell signaling, and the goal of the project is to map complex signaling networks. While cell biologists once believed that signals, such as a drug candidate binding to a cell receptor, initiated only one pathway, it is now clear that a chemical stimulus can excite different networks that interact in complex ways. Combinations of ligands can increase complexity even further. The ultimate goal of the experimental work within the AFCS is threefold: first, to catalogue the "parts" (that is, the stimuli and signaling proteins); second, to identify the interactions between these parts; and third, to generate a computational model of signaling within the cell.[61]

The AFCS comprises eight "wet labs" and one bioinformatics lab. Each wet lab measures a distinct aspect of the effect produced by different ligands. The bioinformatics laboratory is responsible for integrating the data produced

by the eight wet labs. The leaders of the AFCS have determined that, to generate reliable output that can be meaningfully compared and aggregated across labs, laboratory inputs (e.g., cell lines) and procedures must be standardized. Much work has gone into such standardization, and the protocols used are publicly available on the Web.[62] In addition, the direction of future experiments is agreed on collaboratively, based on the data generated thus far.[63]

Another novel aspect of the AFCS involves its lack of emphasis, at least thus far, on conventional publication through peer-reviewed, printed scientific journals. Rather, after some internal review, data publication takes place expeditiously on the Web.[64] Moreover, although the AFCS investigators replicate experiments and analyze data further, and publish those reviews on the Web as "research reports"[65]—they have no formal "head start" in terms of this analysis. In this respect, the AFCS is explicitly modeled on the HGP. The lack of emphasis on conventional publication also coheres with the organizational structure of the AFCS. While most lab directors are senior tenured professors who have advanced through the conventional career track for academic scientists, many of the individuals who work in the AFCS laboratories are on a different, and in some respects novel, career track: They are postdoctoral scientists who are not planning on tenure-track appointments. Many of these individuals plan to go into private sector research.

On the other hand, the AFCS is beginning to shift toward a more-conventional, print-publication-oriented approach. Some of the lab heads observe that it may be difficult to get scientists outside the AFCS to pay attention to the data generated by the alliance without using the conventional publication route.[66] Indeed, the AFCS website now emphasizes that scientists who use AFCS data can publish their work in peer-reviewed publications; a few in fact have published such work.[67] Nonetheless, among the AFCS lab directors, there is lingering concern that prestigious print publication venues—*Science*, *Nature*, *Cell*, and the like—may be reluctant to publish papers associated with data available on the Web prior to publication.[68]

Finally, and perhaps most unusually, all participants in the AFCS agreed to disavow intellectual property rights in their research. This agreement to disavow conventional property rights, quite obviously, is contrary to the trends in patenting we have witnessed since passage of the Bayh-Dole Act. Moreover, many of the institutions participating in the AFCS, perhaps most

notably the University of California system but also the University of Texas and the California Institute of Technology, have substantial numbers of patents. It appears that Gilman's Nobel Prize, as well as his stature in the area of cell signaling, enabled him to convince recalcitrant university administrators, particularly at the University of California and the California Institute of Technology, not to interfere in the disavowal of property rights. But even someone of Gilman's stature found the task difficult.[69]

Open and Collaborative Biomedical Research: A Critical Evaluation

This section assesses the extent to which the open and collaborative approach has the potential to produce socially desirable innovation, particularly as compared with more traditional, proprietary approaches.

Open Source Bioinformatics Software. To some extent, the variables involved in the normative evaluation of open source bioinformatics software are similar to those involved in the normative evaluation of open source software generally. Although no technical evaluation of which I am aware has specifically compared open source bioinformatics software with comparable closed source software, the technical superiority of certain types of open source software suggests that, at least in some circumstances, the open source model might yield technically superior bioinformatics software. At a minimum, open source software may be a good alternative for producing an output of reasonable quality at low cost. In addition, academic computational biologists appear to be motivated to contribute to open source software by some of the same incentives that motivate other open source developers: creative pleasure, the desire to solve a specific problem, sense of community obligation, or perhaps even the possibility of consulting revenues. Equally important, the use of open source software, even copyleft software, should not interfere with ordinary scientific kudos incentives. The open source regime allows publication in conventional print journals of articles that either discuss the software or, more likely, discuss biological insights that are gained by using the software (for example, to query a database).

University-based open source production of bioinformatics software is also unlikely to undermine commercial development of such software,

whether on an open or closed source model. Unlike small biopharmaceutical firms, small software firms do not appear to rely on exclusive licenses to university patents.[70] As a consequence, even if we assume that small firms are important for generating or disseminating research inputs widely, there is little reason to believe that open source production of software within the university will undermine the formation of small firms.

The main obstacle to experimentation with open source bioinformatics software appears to be institutional. As currently constituted, the financial interests of open source developers and university TTOs may be at odds. To the extent that open source developers derive money from consulting revenues, they may be reluctant to embrace any university restrictions on the availability of source code to potential customers. Indeed, in the last few years, there have been a few celebrated cases where open source bioinformatics software developers have clashed with universities over questions of ownership and licensing. These have been cases in which the researcher derived substantial consulting revenue from open source distribution of the bioinformatics software.[71]

The question of who should be responsible for determining whether a particular piece of bioinformatics software is open source is a difficult one. Nonetheless, an argument can be made in favor of the approach taken by MIT, Stanford, and the University of Texas: deference to researcher choice. Although open source is not necessarily the best approach for all software projects (for example, any blanket NIH mandate to require open source software in its grants would be unfortunate), bioinformatics software developers are probably well placed to determine whether, in any given case, open source development is the best approach as a scientific matter. Moreover, in the context of software, researcher choices are unlikely to impose significant development-impeding externalities (as they arguably did in the pre-Bayh-Dole era): The software itself can be developed through volunteer labor, and it is difficult to imagine how any form of open source licensing, even copyleft licensing, would undermine important patent rights on biochemical compounds, such as genes, proteins, or small molecule chemical drugs. Deference to researchers might be particularly desirable to the extent that researcher preferences were not unduly biased in favor of open source because of the prospect of consulting revenues. For example, universities might ask for the same percentage of consulting revenues that they currently get from licensing

revenues. This option would, of course, have the corresponding advantage of not biasing universities against open source.[72]

Open and Collaborative Databases. In several respects, an open and collaborative approach to database generation and improvement has value. Not only does the approach allow for comprehensive database annotation, but it also provides an "infrastructure" of freely available scientific information that all researchers, including wet-lab researchers, can use. Unlike software projects, however, database generation can require substantial capital investment. Although sequencing machines and other tools for generating data have decreased in price, it is unrealistic to imagine that tools for generating data can readily be made available to all interested volunteers. Indeed, in the context of the HGP, efficient production of data required streamlining to a core set of five institutions. In sum, although data generation can be open and collaborative, some restrictions on participation, as well as public funding, will generally be necessary.

In contrast with initial data generation, data annotation can hew more closely to the open source model. In many cases, annotators use publicly available computer algorithms to search existing databases for comparable sequences of known function. Thus, just as with software, the major expense associated with annotation is labor. Indeed, the DAS system used by the public genome projects is in many respects less hierarchical than open source software development.

Because of the high cost associated with generating initial data, and the corresponding value associated with such data, public funding of databases probably undermines the ability of private businesses to form around databases. Unlike software, it is unlikely that private database businesses can be built on a services model. The race to sequence the human genome provided something of a natural experiment for testing the services model. Once the public data were available, the primary value added that Celera could provide was service related. Although a significant number of firms and academic institutions subscribed to the Celera database for these services,[73] the availability of the public data placed a ceiling on what Celera could charge. This ceiling was sufficiently low that Celera has largely moved out of the database business and into drug development. Arguably, the challenge from Celera provided the competition necessary for the public project to streamline its

approach. To the extent that Celera's database business failed, we may see fewer challenges of this sort in the future.

Although a strong argument can be made that upstream data should be publicly available, even if such availability requires public funding and undermines private database companies, the case for having such data undermine the availability of patent rights on more downstream improvements generated by data users is weaker. Unlike software patents, patents on drug candidates, and perhaps even downstream research that leads directly to a drug candidate, are unequivocally important in the biopharmaceutical industry. Moreover, while copyleft licensing might be useful for inducing participation in purely volunteer open source projects,[74] it should not be as critical for inducing such participation in projects where the collaborators are publicly funded academics. A statement by prestigious print publications pledging not to discriminate against articles that analyze publicly available data would provide additional incentive: Although data generators and annotators might not have any official head start in submitting those articles, their familiarity with the data should put them in a good position for submitting original analyses. For example, in the case of the Human Genome Project, participants in the project submitted original analyses that were published by *Nature*. Additionally, at least in the long term, it might be appropriate for the biological community to give data generators and annotators publication-type credit for their work, even if the work is placed immediately on the Web and reviewed by peers subsequent to such Web publication rather than prior to it.[75]

Wet-Lab Biology. Thus far, the open and collaborative model's application to wet-lab biology has largely been limited to systems biology. Even this limited application is significant, however. As a scientific matter, there is reason to believe that systems biology will be crucial for certain types of drug innovation, particularly in the context of complex diseases. Moreover, given the limited capabilities of any single lab, collaboration between labs may be necessary for understanding systems biology.

As with initial data generation, the capital costs associated with wet-lab biology are sufficiently high that it probably is inefficient for most wet-lab collaborations to be open to all comers. Indeed, even with a limited number of players, public funding is necessary. Finally, some level of

standardization of wet-lab inputs may be important to generate reliable data, especially if such data are going to be aggregated across laboratories. At a minimum, transparency of the protocols used to generate the data is necessary.

Although it may be inefficient for collaborative work in the wet lab to move beyond a limited set of players, this number of players is still significantly larger than in traditional wet-lab science. In addition, although this has not yet happened, it is certainly possible that annotation of the data generated by a wet-lab collaboration could invoke the DAS model and thus encompass a much larger group.

For collaborative projects, the gains that can accrue from disavowal of intellectual property rights are significant. Unlike the AFCS, universities and investigators involved in other large-scale collaborative projects funded by the NIGMS have not similarly disavowed such rights. Without disavowal of intellectual property rights, concerns that information exchange will lead to public disclosure of proprietary information, as well as disputes over how to allocate patent rights among a host of potential university assignees, may create friction. The principal investigator of one consortium that has not disavowed proprietary rights, Alan Horwitz of the Cell Migration Consortium, reports some dissatisfaction with the manner in which the relevant university TTOs in his consortium conducted their negotiations. He believes that TTO-imposed requirements whereby each university agrees to keep strictly confidential and refrain from commercializing the proprietary information of other universities in the consortium have "gotten in the way of the science."[76] Horwitz hopes that the relevant interuniversity agreements will be renegotiated in the future. The possibility that better agreements will be produced in the future is not necessarily high, however. Within wet-lab biomedical research, universities jealously guard their ability to commercialize proprietary information, particularly by turning it into patents. To turn proprietary information into patents, strict restrictions on dissemination are necessary: The relevant court decisions by the Federal Circuit hold that even limited public sharing of information can create patent-defeating prior art.[77] The prevailing regime governing exploitation rules for patents also makes collaborative research difficult. For example, each inventor or assignee to whom the right has been assigned can fully exploit the patent and there is no duty to account. As Rochelle Dreyfuss

has noted, "The result of these exploitation rules is a rivalry that is potentially so destructive [that] the need to consolidate rights in a single owner is overwhelming."[78]

The disavowal of intellectual property rights in wet-lab biology raises concerns about commercialization. It is certainly possible that some of the information placed in the public domain by projects like the AFCS will be left to linger in an undeveloped state, as feared by the proponents of Bayh-Dole. At least in the specific context of the AFCS, however, the fact that pharmaceutical companies fund some of the research indicates that commercialization is unlikely to be altogether defeated. More generally, the evidence that public domain status for upstream information defeats commercialization is hardly solid. What public domain status for such information may do is undermine some small biotechnology firms. However, small biotechnology firms are unlikely to have the resources necessary to conduct large-scale systems biology experiments.

As an institutional matter, if projects like the AFCS are going to work, either their publication model or the print publication emphasis of the biological sciences may need modification. As matters currently stand, the lack of emphasis on print publication, coupled with the disavowal of patents, has meant that only the AFCS lab heads are traditional academics. To the extent that complex systems biology projects will succeed only if they attract the most academically oriented young minds, the failure to attract such researchers is worrisome. It is important, therefore, that the AFCS appear to be moving toward more conventional publication for its own investigators. It is also important that the AFCS is explicitly encouraging other investigators to use its data as the basis for publication in peer-reviewed journals. As noted earlier, a statement by prestigious peer-reviewed publications making it clear they do not discriminate against articles based on data already made publicly available would be useful. In the long term, a move in the biological sciences toward a model that emphasizes (or at least recognizes) Web-based publication, with subsequent peer review, is also worth considering.

From an institutional standpoint, it is also worth noting that university agreements to disavow intellectual property rights are not easy to achieve. Only charismatic and superbly credentialed scientists like Al Gilman are likely to secure such agreements. They do so by convincing their scientific

colleagues that they must sign on to a particular research project, whatever the political difficulties. Moreover, even individuals like Gilman probably have to be supported by the relevant funding agency.

Given these institutional obstacles, one might reasonably ask whether the large-scale collaborative work necessary for systems biology needs to be done using consortia of the type represented by the AFCS. Rather, a large pharmaceutical firm might be the appropriate arena for such work. One might imagine, for example, that Pfizer could set up eight laboratories like the AFCS laboratories. To guard against the innovation-suppressing tendencies of large hierarchies, these laboratories could also be given substantial freedom to pursue the projects that they want to pursue. A Pfizer-based collaborative project would not have to worry about allocating intellectual property rights. Moreover, given that all work is done within a single firm, it is unlikely that even robust exchange of information within the firm would be deemed public, so as to create patent-defeating prior art.

Interestingly, there is some movement in the pharmaceutical industry toward large-scale collaboration. However, the collaboration contemplated is not within the firm but *between* firms. An organization called the CEO Roundtable on Cancer is considering a proposal to create a research collaboration among a large number of companies for purposes of making an "all-out effort against cancer." Various pharmaceutical firms are considering this collaboration between firms, even though the obstacles related to allocation of intellectual property rights—not to mention antitrust concerns—are quite considerable.[79] The cost to a single company of doing such systems biology research may be sufficiently great or the possibility of attaching intellectual property rights that recoup the cost sufficiently small that no single company wants to engage in the research. In that case, the research may be better produced as a public good, as by the AFCS.

Conclusion

Approaches to biomedical research in which such research is generated and improved upon in an open, collaborative fashion represent a potentially valuable experiment. The most intuitive case for such an experiment is open source bioinformatics software. In the case of software, the major

obstacle to successful experimentation could be removed by instituting contractual mechanisms to divide consulting revenues between investigators and universities. With respect to open and collaborative databases, the argument is somewhat more equivocal. Nonetheless, when the data in question are upstream, a significant case can be made in favor of publicly funded, publicly available databases that can be improved on collaboratively. The case becomes weaker as the information being produced is downstream in the research path. Rather than using copyleft style licensing that undermines patents on downstream information, it may be advisable to attract collaborators by attempting to change biological science norms regarding publication. Such norm change would also improve the value of experimentation with large-scale collaboration in wet-lab systems biology. In particular, it would help such collaborations to attract promising young investigators. Although open and collaborative research represents a paradigm shift for wet-lab biology, experimentation with such a paradigm shift might be necessary for solving the intractable biological problems that are currently impeding the development of breakthrough drugs. Moreover, to the extent that large-scale wet-lab collaborations that disavow upstream intellectual property rights can find pharmaceutical company support, they are unlikely to undermine critical patents.

Notes

1. See, e.g., John P. Walsh, Ashish Arora, and Wesley M. Cohen, "Effects of Research Tool Patents and Licensing on Biomedical Innovation," in *Patents in the Knowledge-Based Economy*, ed. Wesley M. Cohen and Stephen A. Merrill (Washington, D.C.: National Academies Press, 2003), 285 (discussing increased concentration of proprietary rights in upstream biomedical research).

2. See, e.g., Eric G. Campbell et al., "Data Withholding in Academic Genetics: Data from a National Survey," *JAMA* 287 (2002): 473; John Walsh and Wei Hong, "Secrecy Is Increasing in Step with Competition," *Nature* 422 (2003): 801 (empirical findings indicating increased secrecy in biomedical research). I discuss the extent to which increases in proprietary rights and secrecy appear to be correlated later in the chapter.

3. For examples of such calls for access, see, e.g., National Research Council, *Sharing Publication-related Data and Materials: Responsibilities of Authorship in the Life Sciences* (Washington, D.C.: National Academies Press, 2003); Department of Health and Human Services, National Institutes of Health, "Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice," *Federal Register* 64 (December 23, 1999): 72,090, 72,093 (research tools).

4. The term *open and collaborative projects* was recently invoked in a letter to the World Intellectual Property Organization (WIPO) urging WIPO to hold an exploratory meeting on these types of projects. See "Letter from Sixty-eight Scientists and Economists to Kamil Idris, Director General of the World Intellectual Property Organization, July 7, 2003," available at www.cptech.org/ip/wipo/kamil-idris-7July2003.pdf. That letter does not specifically define its use of the term. This chapter's definition is set out in the text.

5. See, e.g., William Jorgensen, "The Many Roles of Computation in Drug Discovery," *Science* 303 (2004): 1818.

6. See International HapMap Project Public Access License, available at www.hapmap.org/cgi-perl/registration (acknowledging model of GNU General Public License).

7. See, e.g., Dan Burk, "Open Source Genomics," *Boston University Journal of Science and Technology* 8 (2002): 254, 255. For a preliminary empirical investigation in the journalistic literature, see Kenneth Neil Cukier, "Community Property: Open-Source Proponents Plant the Seeds of New Landscape," *Acumen* 1 (2003): 57–58.

8. For purposes of this paper, I focus on research that addresses diseases prevalent in developed countries and assume (somewhat heroically) that market demand accurately reflects the socially desirable direction and rate of innovation.

9. Caitlin Smith, "A Question of Biology," *Nature* 428 (2004): 225, 231 (noting comment by scientist that "[i]n the past, a simplistic view was, by necessity, taken,

which resulted in many drugs failing in preclinical or clinical trials for lack of efficacy or side effects. The new approach must account for this complexity. The holistic systems biology approach to research will be necessary to overcome this challenge.").

10. Iain M. Cockburn, "The Changing Structure of the Pharmaceutical Industry," *Health Affairs* 23 (2004): 10, 11; Robert F. Service, "Surviving the Blockbuster Syndrome," *Science* 303 (2004): 1796 (discussing low number of new chemical entities approved in 2002).

11. See, e.g., Wesley Cohen et al., "Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)" (working paper no. 7552, NBER, Cambridge, Mass., 2000).

12. Compare Robert K. Merton, "The Normative Structure of Science," in *The Sociology of Science* (Chicago: University of Chicago Press, 1973) (famously arguing that, in general, academic research science is a communal enterprise).

13. Walsh and Hong, "Secrecy Is Increasing." Cited in note 2.

14. See, e.g., Arti K. Rai, "Regulating Scientific Research: Intellectual Property Rights and the Norms of Science," *Northwestern Law Review* 94 (1999): 77, 92.

15. The view of patents enunciated in Bayh-Dole is closely associated with those scholars who view patents as directly analogous to rights in tangible property. Like the grant of rights in tangible property, an early decision to grant patent rights ensures that the subject of the right is developed or utilized at the appropriate rate, neither "too fast" nor "too slow." See Edmund Kitch, "The Nature and Function of the Patent System," *Journal of Law and Economics* 20 (1977): 265 (applying property rights theory enunciated by Harold Demsetz to patents and discussing how an early grant of rights can guard against both "racing" and underdevelopment).

16. See, e.g., *In re Brana*, 53 F.3d 1560 (Fed.Cir. 1995) (utility); *In re Deuel*, 51 F.3d 1552 (Fed. Cir. 1995) (nonobviousness).

17. International Patent Classification categories, coupled with data on all university patents, can provide some rough sense of the university patent presence in biomedical research. In 2000, the percentage of biomedical research patents secured by research universities was somewhere between 11 percent (using a "liberal" definition of biomedical research that may sweep in some patents on end products) to 15 percent (using a more "conservative" definition of biomedical research). This compares with a university percentage of between 2 to 5 percent in 1980. Thanks to Bhaven Sampat (assistant professor, School of Public Policy, Georgia Institute of Technology) for this data.

18. See, e.g., *University of Rochester v. G. D. Searle*, 358 F.2d 916 (2004); *Regents of the Univ. of Cal. v. Eli Lilly*, 119 F.3d 1559 (Fed. Cir. 1997) (striking down broad patents on biomedical research).

19. *Amgen v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313 (upholding broad biomedical patent).

20. Empirical evidence regarding the superior innovative capacities of small firms is mixed, however. See Zoltan J. Acs and David B. Audretsch, "Innovation in

Large and Small Firms: An Empirical Analysis," *American Economic Review* 78 (1988): 678 (finding that in twenty-one of thirty-five industries, large firms were more innovative than small firms).

21. Ronald Coase, "A Theory of the Firm," *Economica* 4 (1937): 386.

22. Walsh, Arora, and Cohen, "Effects of Research Tool Patents." Cited in note 1.

23. Ibid.; see also Josh Lerner, "Patenting in the Shadow of Competitors," *Journal of Law and Economics* 38 (1995): 463.

24. *Rochester v. Searle; Ariad v. Eli Lilly.*

25. Campbell et al., "Data Withholding." Cited in note 2. Thirty-five percent of geneticists said that sharing had decreased during the 1990s, whereas only 14 percent said that sharing had increased.

26. See 35 U.S.C. 102(b) (establishing that an invention cannot be patented if it has been disclosed publicly more than one year before a patent application is filed).

27. See David Blumenthal et al., "Withholding Research Results in Academic Life Science," *JAMA* 277 (1997): 1224.

28. Campbell et al., "Data Withholding." Cited in note 2.

29. Ibid. Cf. John P. Walsh, "For Money or Glory? Secrecy, Competition, and Commercialization in Science" (presentation at American Sociological Association Annual Meeting, August 16, 2004) (finding increased secrecy but determining it was not associated with a patent application).

30. Cf. Elinor Ostrom, *Governing the Commons* (Cambridge: Cambridge University Press, 1990), 88–89 (noting that enduring norms operate where the group is small and has similar interests and values).

31. Bruce Perens, "The Open Source Definition," available at http://perens.com/articles/osd.html.

32. Although Richard Stallman and some others argue that copylefted software should be called free software, this paper uses the term *open source* to encompass copylefted software.

33. Jonathan Zittrain, "Evaluating Free and Proprietary Software," *University of Chicago Law Review* 71 (2004): 265, 279. According to Josh Lerner and Jean Tirole, approximately 70 percent of the 25,729 projects found at www.sourceforge.net used GPL-style licenses. Joshua Lerner and Jean Tirole, "The Scope of Open Source Licensing," (Harvard NOM working paper no. 02-42, 2002), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=354220.

34. Discussion at American Bar Association, Joint Session of Intellectual Property Section and Science and Technology Section, April 1, 2004. There are two recorded instances of litigation brought by holders of copyleft licenses claiming improper propertization of the code. Zittrain, "Evaluating Free and Proprietary Software," 285. Cited in note 33.

35. S. O'Mahony, "Guarding the Commons: How Community Managed Projects Protect Their Work," *Research Policy* 32 (2003): 1179, 1189.

36. See generally Eric von Hippel and Georg von Krogh, "Editorial, Special Issue of Open Source Software Development," *Research Policy* 32 (2003): 1149. The central developers' control of the project is sufficiently high that "forking" of the source code is rare. Eric Raymond, "The Magic Cauldron," sections 3–5, available at http://www.catb.org/~ESR/writings/magic-cauldron/magic-cauldron.html.

37. Compare David McGowan, "Legal Implications of Open-Source Software," *University of Illinois Law Review* (2001): 241 (discussing respects in which open source software production is and is not like firm-based production).

38. According to one survey, only about 7 percent of open source software developers work in the academic sector. Karim R. Lakhani and Robert G. Wolf, "Why Hackers Do What They Do: Understanding Motivation Effort in Free/Open Source Software Projects," in *Perspectives on Free and Open Source Software*, ed. J. Feller, B. Fitzgerald, S. Hissam, and K. R. Lakhani (Cambridge, Mass.: MIT Press, 2005).

39. See, generally, ibid.; see also Lerner and Tirole, "Scope of Open Source Licensing." Cited in note 33.

40. James W. Paulson et al., "An Empirical Study of Open-Source and Closed-Source Software Products," *IEEE Transactions on Software Engineering* SE-30 (2004): 246.

41. For firms that use software platforms, the cost of the software may be only a small part of the total cost of ownership. In particular, staffing costs for such platforms can be quite high. See Alan MacCormack, "Evaluating Total Cost of Ownership for Software Platforms: Comparing Apples, Oranges, and Cucumbers," available at http://www.aei.brookings.org/admin/authorpdfs/page.php?id=261. Staffing costs are likely to be less important, however, where the software in question is itself directed at a technical audience (as is most bioinformatics software).

42. Yochai Benkler, "Coase's Penguin, or Linux and the Nature of the Firm," *Yale Law Journal* 112 (2002): 369.

43. Interview with computational biologist Steven Brenner, UC Berkeley, March 13, 2004. According to the founding developers of Bioperl, "[a] primary motivation behind writing the toolkit is the authors' desire to focus energies on a solution whose components can be shared rather than duplicating effort." Jason E. Stajich et al., "The Bioperl Toolkit: Perl Modules for the Life Sciences," *Genome Research* 12 (2002): 1611.

44. Ibid. See also Russ Altman et al., "Whitepaper on Open Source Software in Bioinformatics," available at https://www.iscb.org/lists/iscb-software.html/dir/msg0000.shtml (on file with author) (arguing that NIH should not mandate open source for its grant recipients).

45. See Arti Rai, "Open and Collaborative Biomedical Research: An Empirical Look" (draft paper, on file with author) (discussing policies of twenty research universities with large software patent, biomedical patent, and/or biomedical research portfolios).

46. Ibid.

47. University representations are borne out by the numbers. One estimate based on International Patent Classifications (IPCs) common for software publishing industry patents indicates that universities actually patented less as a relative matter in 2000 than they did in 1980. While university software patents represented 1 percent of all software in 1980, they represented 0.6 percent of all such patents in 2000. Thanks to Bhaven Sampat for data on software patents.

48. Rai, "Open and Collaborative Biomedical Research." Cited in note 45.

49. Ibid.

50. Interviews with Lita Nelsen (director, Technology Licensing Office, MIT) and Kathy Ku (director, Stanford Office of Technology and Licensing), September 5, 2003.

51. Interview with Georgia Harper (manager, Intellectual Property Section, University of Texas Office of General Counsel), February 5, 2003.

52. In January 2003, NHGRI extended this policy of immediate data deposition without accompanying intellectual property rights to all large-scale data "infrastructure" projects. Indeed, at this meeting, NHGRI gave higher priority to immediate and full access to data than to the traditional scientific norm that the investigator who generates the data has the right to do the first analysis of this data. *Nature* 421 (2003): 875.

53. Interview with Lincoln Stein, March 26, 2004.

54. See Eliot Marshall, "Genome Researchers Take the Pledge: Data Sharing," *Science*, April 26, 1996, 478 (noting that key university patent officials approved of policy). One leading officer, Lita Nelsen of MIT, noted, however, that she is wary of the "bad precedent" that the policy might set. Ibid.

55. John Sulston, *The Common Thread* (Washington, D.C.: Joseph Henry Press, 2002), 211.

56. Similarly, another important open and collaborative project that took place at approximately the same time as the HGP, the Single Nucleotide Polymorphism (SNP) Consortium, put its data in the public domain.

57. See International HapMap Project Public Access License, available at www.hapmap.org/cgi-perl/registration.

58. Because there is no underlying copyright, those who manage to access the data without having agreed to the license are not subject to any legal prohibition against patenting. The relative weakness of the HapMap prohibition is probably salutary, however, because, as discussed later in the chapter, a direct application of copyleft licensing for biological databases may impede commercialization unduly. On December 10, 2004, as this chapter was going to press, the leaders of the HapMap project announced that license restrictions had been lifted because haplotype information was now available.

59. The creation of genome sequence databases also has a wet-lab component. The wet-lab component was particularly substantial before the widespread dissemination of automated laser-based sequencing machines. But the wet-lab component is small and relatively easy to standardize, at least as compared to traditional biology.

60. Interview with Al Gilman, September 4, 2003.

61. Interview with Shankar Subramanian (professor of bioengineering, UCSD), March 17, 2004.

62. See www.signaling-gateway.org/data/Protocol/Links.html.

63. Interview with Alex Brown (Ingram Associate Professor of Cancer Research, Vanderbilt University), April 23, 2004.

64. See www.signaling-gateway.org/data/Data.html (AFCS data center, hosted by AFCS and *Nature*).

65. See www.signaling-gateway.org/reports/ReportCover.html.

66. Interview with Alex Brown, March 16, 2004. Brown notes that most data, including the data generated by the AFCS, are not as publicly visible, as were the data from the Human Genome Project.

67. See www.signaling-gateway.org/reports/JournalPubs.htm.

68. Interview with Alex Brown, April 23, 2004.

69. Interview with Al Gilman, September 4, 2003.

70. More generally, the role played by patents in the software industry is unclear. In contrast, drug patents, and perhaps even patents on certain types of upstream work, are clearly important to the biopharmaceutical industry.

71. Memorandum from Pat Jones (director, Office of Technology Transfer, University of Arizona) discussing cases, March 2004.

72. One concern that has recently emerged is the possibility of contributors to open source adding code that may have property rights attached to it. Because it is very difficult for open source project leaders to verify that contributors are adding code that is free of proprietary rights, the *SCO v. IBM* lawsuit, in which SCO claims copyright interests over parts of UNIX that have allegedly been incorporated into Linux, has generated much concern in the open source community. At this early stage, however, the potential implications of this lawsuit, particularly for university-based open source researchers, are difficult to gauge.

73. James Shreeve, *Genome Wars* (New York: Knopf, 2004), 368–69.

74. See supra note 34.

75. This model has been used successfully, for example, in other natural sciences, primarily physics. See www.arXiv.org.

76. Interview with Alan Horwitz, March 31, 2004.

77. The recently enacted Cooperative Research and Technology Enhancement Act of 2004 (CREATE) aims to encourage collaborations by reducing the likelihood that so-called secret prior art created by the collaboration will defeat patentability. But this law does not address public disclosure of prior art created by the collaboration.

78. Rochelle Dreyfuss, "Collaborative Research: Conflicts on Authorship, Ownership, and Accountability," *Vanderbilt Law Review* 53 (2000): 1161.

79. Susan Warner, "Collaborators against Cancer," *Scientist*, June 30, 2003 (noting these obstacles).